

Singular Spectrum Analysis를 이용 필터링된 자료로부터의 엘니뇨 빈도분석

문 영일¹⁾, ○박 대형²⁾, 황 성환³⁾

1. 서론

전세계적으로 이상 기상현상에 영향을 주었던 97/98년 엘니뇨는 97년 3월부터 시작하여 그 위세가 굉장히 강해져 관측 이래 가장 강한 엘니뇨 중의 하나로 기록되어 세계의 이목을 집중시켰다. 1998년 세계 기상기구에 의하면 1997년은 세계 평균기온이 평년보다 0.44°C 높아 1950년 이래 기온이 가장 높았던 해로 기록되어 엘니뇨가 기온상승에 영향을 주고 있음을 보여 주었다. 강한 엘니뇨의 발달에 따라 태평양에서 대기중으로 방출된 에너지는 전세계적인 이상기후를 초래하였다. 1998년 여름철 이전까지는 기상재해가 열대지방에 집중되었으나, 여름철부터는 양쯔강 홍수, 우리나라의 집중호우, 일본의 홍수 등 아시아지역에서 예측하기 어려운 심각한 기상재해가 발생하였다.

이런 엘니뇨현상은 지역적인 의미로는 동태평양 페루 연안의 이례적인 해수면 온도 상승을 지칭하는데, 통상적으로 동태평양 페루 연안에는 용승현상으로 인한 차가운 하층해수가 해수면으로 용출함에 따라 저온의 해수면 온도를 유지하게 되지만, 어느 해 용승효과가 약화되고 더운 해수가 동쪽으로 이동하게 되면 중부와 동부 적도 태평양의 해수면온도는 점차 상승하게 되는 현상이다. 그러므로 이 현상은 페루 연안에서만 발생하는 국지적인 현상이 아니라 페루 연안에서 날짜 변경선까지 약 1만km에 달하는 지역에서 발생하는 큰 규모의 현상이며, 1997-98년 엘니뇨의 경우에는 위성자료의 분석결과 지구의 1/3에 해당하는 지역이 엘니뇨의 영향권내에 있었음을 보여주고 있다.

이런 엘니뇨현상을 나타내는 가장 기본적 자료는 해양성분인 해수면온도(Sea Surface Temperature, SST)와 대기성분인 해수면 압력(Sea Level Pressure, SLP)이다. 현재 SST는 Nino1+2, Nino3, Nino4, Nino3.5, Ship track1, 그리고 Ship track6의 6개 지역에서 관측되고 있다. 이 지역 중에서 엘니뇨현상과 가장 강한 상관성을 보이는 곳은 Nino3.5지역이며, Trenberth(1997)에 의하면 이 지역에서 관측된 SST 값의 5개월 이동평균이 6개월 이상 0.4°C를 초과할 때를 엘니뇨현상으로 구분하고 있다. 또한 엘니뇨현상으로 인하여 남동태평양과 남서태평양 사이에 남방진동(Southern Oscillation)이라 알려진 압력의 변화현상이 반복적으로 일어나는 것이 관측되었다. 즉, 남동태평양상에는 고압현상이, 남서태평양상에는 저압 현상이, 또는 반대로 남동태평양상에서는 저압현상이 남서태평양에서는 고압현상이 지속적으로 나타난다. 이러한 현상을 나타내는 정량화된 지수가 남방진동지수(Southern Oscillation Index, SOI)이다. 남방진동지수는 엘니뇨현상의 크기를 나타내 주는 지표로 남방진동지수가 큰 음수값을 가지면 강한 엘니뇨사상이 진행중이고, 반대로 큰 양수값을 가지면 강한 라니냐(엘니뇨의 반대현상)가 진행 중인 것을 나타낸다. 남방진동지수 SOI값은 남태평양상에 있는 두 도시, Tahiti(150 °W, 17.5 °S)와 Darwin(131 °W, 12.5 °S)간의 평균해수면 압력(SLP)의 차이 값이다. Kiladis와 Diaz(1989)는 SOI의 값이 3개월 이상 -1.0 이하일 때를 엘니뇨현상으로 구분하였다.

엘니뇨에 대한 최근의 연구는 공간-시간 접근보다는 SOI 자료를 이용한 시간 영역에서의 연구가 활발해지는 경향이 있다. 즉, 비교적 일반적으로 비슷한 지역에서 발생하는 경향이 있는 엘니뇨현상의 공간적 분포보다는 발생시기를 규명하고자하는 기상학적 변동성에 대한 시간분포

1) 서울시립대학교 토목공학과 조교수
2) 서울시립대학교 토목공학과 석사과정
3) 서울시립대학교 토목공학과 석사과정

에 대해 연구가 지속되고 있다(Barnett 등, 1988; Kiladis와 Diaz, 1989; Ghil과 Vautard, 1991; Keppenne과 Ghil,1992; Moon과 Lall, 1996). 즉, 미래에 발생할 엘니뇨의 시기를 예측하여 발생될 피해를 최대로 줄이기 위해 엘니뇨 예측모형에 대한 연구가 활발하게 지속되고 있다. 여기서 앞으로 발생할 엘니뇨의 시기뿐만 아니라 얼마만한 크기의 엘니뇨가 발생할 것인가 또한 매우 중요한 인자이다. 따라서 본 연구에서는 NOAA(National Oceanic and Atmospheric Administration in USA)에서 제공한 1932년 9월부터 1999년 3월까지의 799개의 월별 SOI자료를 기초로 하여 과거 관측된 엘니뇨현상들을 빈도해석을 통하여 재현기간별로 분류하여 보았다. 직접 SOI 자료를 이용한 빈도해석보다 SOI 자료에 SSA(Singular Spectrum Analysis)를 적용하여 자료를 필터링한 후, 재구성 된 자료를 이용하여 자료자체의 특성을 잘 표현해 주는 비매개변수적 변동 핵밀도함수를 이용하여 빈도분석을 하였다.

2. Singular Spectrum Analysis(SSA)

Singular Spectrum Analysis(SSA)는 주성분분석(Principal Component Analysis, PCA)에서 파생된 기법으로, 기본적인 개념은 자료의 비조화성분을 제거함으로써 원자료에서 얻지 못하는 주기성과 경향성(trend)을 쉽게 파악할 수 있는 방법이다(Moon과 Lall, 1996). 즉, SSA는 주어진 자료의 시계열을 여러 개의 조화성분으로 나눌 수 있으므로 자료가 갖고 있는 경향을 파악할 수 있다. 이러한 자료의 경향성을 통하여 앞으로 발생할 사상에 대해 예측하는 능력 또한 다른 기법들에 비해 뛰어나다. 이 SSA 방법은 Colebrook(1978)에 의해 해양생물학에 처음으로 사용되었으며, 80년대의 Broomhead와 King(1986), Fraedrich(1986)을 거쳐 90년대의 Vautard와 Ghil(1989), Rasmusson(1990), Vautard 등(1992), Keppenne과 Ghil(1992), Moon과 Lall(1996)등이 엘니뇨현상을 연구하기 위해 사용하였다.

만일 시계열의 값이 x_i ($1 \leq i \leq N$)라고 하면 아래와 같이 표현할 수 있다.

$$x_{i+(j-1)\tau} = \sum_{k=1}^M a_i^k E_j^k, \quad 1 \leq j \leq M, \quad 1 \leq i \leq N-M+1 \quad (1)$$

여기서 a_i^k 는 주요성분(Principal Component, PC)이라고 불리는 투영계수이다. E^k 는 경험적 직각함수(Empirical Orthogonal Function, EOF) ($1 \leq k \leq M$)이며 M은 M차의 공간 (smoothing window의 역할)을 의미하며 τ 는 추출비(sampling rate)이다. k번째 PC는 k번째 EOF에 대하여 시계열의 정투영(orthogonal projection)으로 정의된다.

$$a_i^k = \sum_{j=1}^M x_{i+(j-1)\tau+1} E_j^k \quad 0 \leq i \leq N-M \quad (2)$$

따라서 고유치(eigenvalue)의 크기에 따라 자료의 비조화성분을 분리할 수 있고 아래와 같이 원 자료를 재구성할 수 있으며 이를 재구성성분(Reconstructed Component, RC)이라 한다. 따라서 M개의 RC들을 전부 합하면 원 시계열을 나타낸다.

$$(R_A x)_i = \frac{1}{i} \sum_{j=1}^i \sum_{k \in A} a_{i-j}^k E_j^k, \quad 1 \leq i \leq M-1 \quad (3)$$

$$(R_A x)_i = \frac{1}{M} \sum_{j=1}^M \sum_{k \in A} a_{i-j}^k E_j^k, \quad M \leq i \leq N-M+1 \quad (4)$$

$$(R_A x)_i = \frac{1}{N-i+1} \sum_{j=1}^M \sum_{k \in A} a_{i-j}^k E_j^k, \quad N-M+2 \leq i \leq N \quad (5)$$

본 논문은 1932년 9월부터 1999년 3월까지 799개의 월별 SOI 자료에 M=60, $\tau=1$ 을 사용하였다. SSA를 통하여 발생한 60개의 고유치(eigenvalue)를 총 분산에 대한 백분율로 표시한 값이 표 1과 같이 나타났다. 표1에서와 같이 상위 6개의 고유치가 총 분산의 60%를 나타내고 있음을 알 수 있다. 그림1에서는 대응하는 고유치에 따라 선택된 세 개의 재구성 성분(RC1-2, RC3-4, RC5-6)을 나타내고 있다. RC1-2는 약 5년의 주기의 SOI 성분을 나타내고, RC3-4는 약 3.5년의 주기, RC5-6은 약 2년의 주기를 나타내고 있다. 즉, RC1-6은 엘니뇨의 저빈도(low frequency) 성분을 나타내고 있으며, 2년 이하의 고빈도(high frequency) 성분은 SOI 자료에서

표 1. 총분산에 대한 백분율로 표시된 고유치(eigenvalue).

고유치	1	2	3	4	5	6	7-60
백분율(%)	13	13	10	10	8	6	40

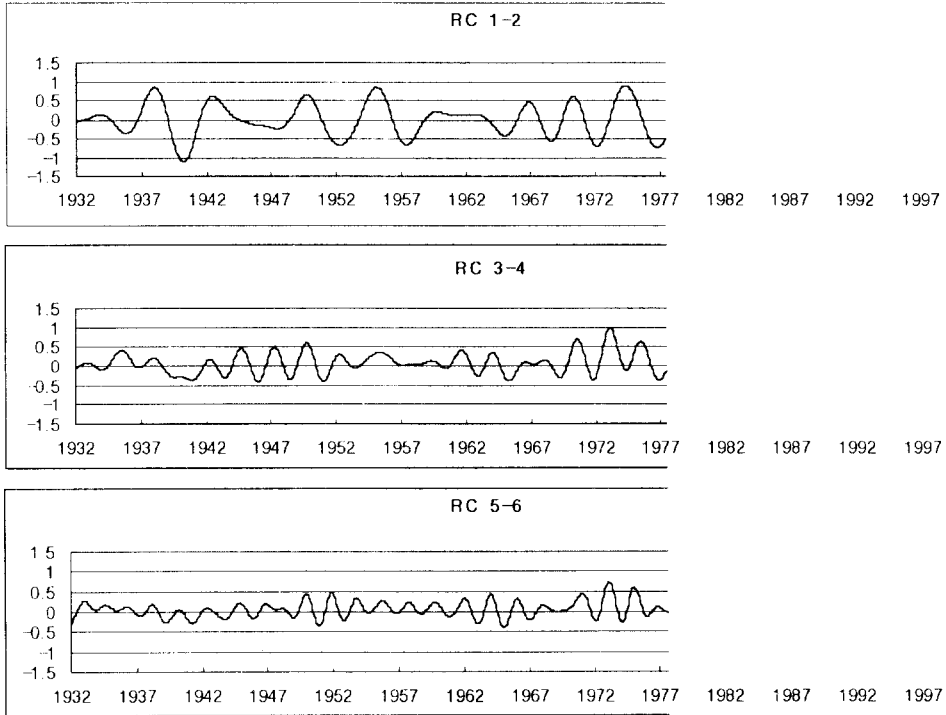


그림 1. 그림 3 SSA를 통하여 재구성된 RCs

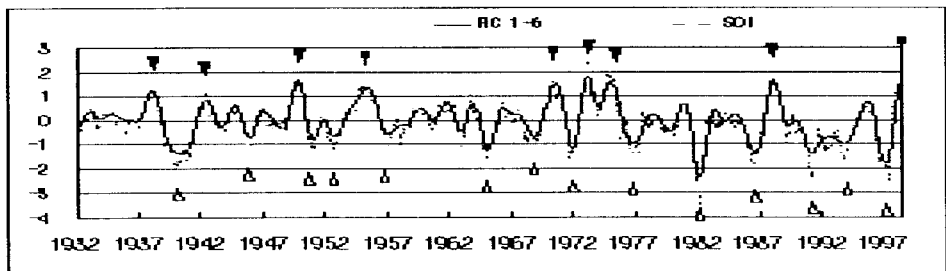


그림 2. SOI 와 RC 1-6과의 비교.

필터링 하였다.

이렇게 SOI 자료를 필터링하여 재구성한 RC1-6을 SOI 자료와 비교하기 위해서 그림2에 도시하였다. SOI 자료는 미국 National Meteorological Center에서 추천하는 5개월 이동평균(moving average)을 사용하였으며(Keppenne과 Ghil, 1992), 이 방법은 엘니뇨현상의 비조화성분을 제거하고 주기를 잘 표현해 주는 것으로 알려지고 있다. 그림2에서 RC1-6과 SOI의 상관

계수는 0.95이며, RC1-6은 엘니뇨 및 라니냐의 발생시기를 잘 표현 해 주고 있다. 그림 2에 도시된 하얀 색의 빈 화살표는 엘니뇨현상이 발생된 때를 나타내며 빨간 색의 화살표는 라니냐현상이 발생한 때를 표시한다

3. 비매개변수적 변동 핵밀도함수(Nonparametric Variable Kernel Density Function)

임의의 자료에 대한 분포함수의 해석방법으로 가장 오래되고 가장 쉽게 사용되는 확률밀도함수는 아마도 히스토그램일 것이다. 이 히스토그램의 확률밀도함수의 식은 다음과 같이 정의할 수 있다.

$$f(x) = \frac{1}{nh} \quad (x \text{가 포함된 계급구간에서의 자료개수}) \quad (6)$$

여기서, n은 자료의 총개수이며 h는 구간의 폭이다. 그러나 히스토그램의 확률밀도함수는 자료의 계급구간이 변화는 점에서 불연속적이고, 구간간격과 시작점의 선택에 따라 확률밀도함수의 모양이 변하기 때문에 사용하기에 어려움이 따른다. 이와 같은 단점을 보완하여 자료의 각 관측치 위치에 히스토그램의 box 대신 연속함수인 핵함수(kernel function)를 아래 그림3과 같이 위치하여 자료의 계급구간이 변화는 점에서 연속적이며 구간을 이동시킬 수 있는 히스토그램이 핵밀도함수(kernel density function)이다. 그림3에서는 정규분포 핵함수와 bandwidth h가 0.5를 사용하였다.

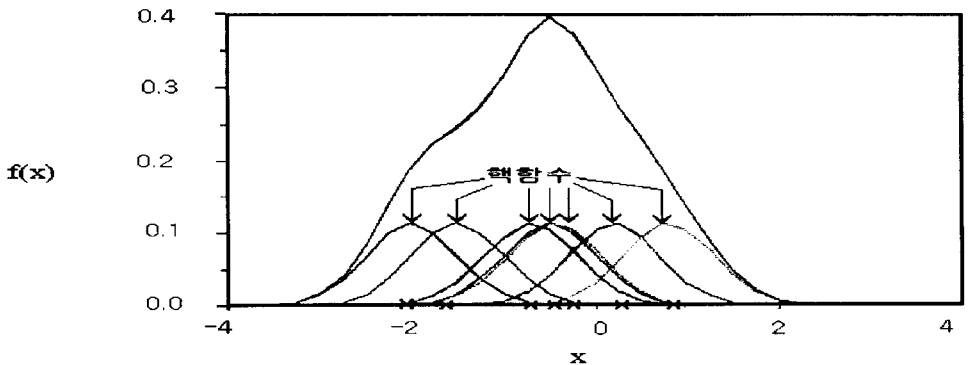


그림 3. 히스토그램의 box 대신 핵함수를 사용한 핵밀도함수.

위와 같은 핵밀도함수의 추정식은 모든 실수 x에 대하여 다음과 같이 정의할 수 있다 (Rosenblatt, 1956).

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (7)$$

여기서 X_1, X_2, \dots, X_n 은 독립적으로 동일하게 분포된 실관측치이다. $K(\cdot)$ 는 핵함수이고 h는 n이 무한대로 갈 때 0으로 접근하는 값을 갖는 양의 bandwidth이다. 이 때 h가 자료의 상황에 따라 변하면 이를 변동 핵밀도함수(variable density function)라 한다. 더욱 상세한 핵밀도함수 추정에 관한 내용은 Silverman(1986), Scott(1992), Lall 등 (1993), Moon 등(1993)을 참조할 것은 권한다.

4. 결과 및 결론

먼저 빈도해석을 위하여 RC1-6과 SOI 자료로부터 각 연도별 최대치계열(라니냐 사상)과 최

소치계열(엘니뇨 사상)을 구성하였다. 이 자료들의 확률밀도함수(probability density function)가 그림4에서 엘니뇨 사상과 라니냐 사상에 대하여 각각 나타나 있다. RC1-6 자료의 경우 라니냐 사상의 확률밀도함수는 bimodal을 가지고 있으며 엘니뇨 사상은 다소 왜곡되게 나타났다. SOI 자료에 대하여는 엘니뇨 사상의 확률밀도함수가 bimodal 형태를 보이고 있다. 이러한 bimodal 형태의 자료에 대응하여 본래 자료의 특성을 잘 표현하는 비매개변수적 변동 핵밀도함수를 사용하여 빈도해석을 하였다. 변동 핵밀도함수에 의한 빈도해석을 통하여 나타난 엘니뇨의 사상별 재현기간이 SOI 자료를 사용하였을 때와 RC1-6을 자료로 사용하였을 때에 각각 재현기간을 추정하여 비교하였으며 이에 대한 결과를 표2에 나타내었다.

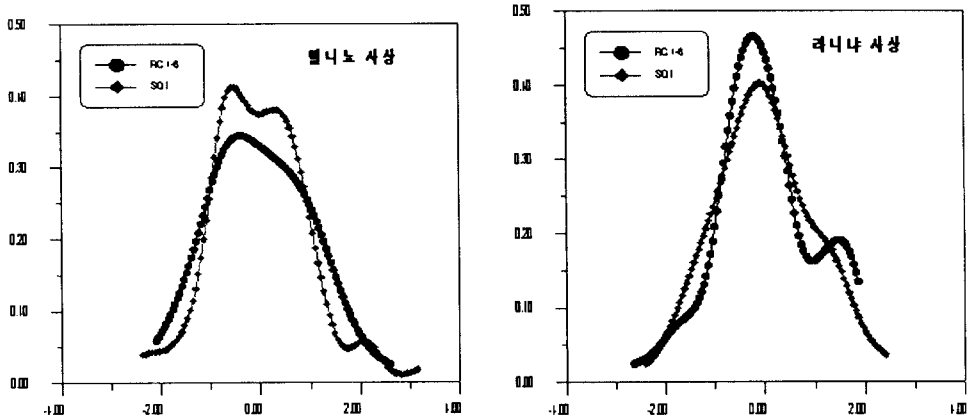


그림 4. SOI와 RC1-6 자료를 이용한 엘니뇨 사상과 라니냐 사상의 확률밀도함수.

표 2. SOI 자료 및 RC 1-6의 자료를 사용할 때의 재현기간.

엘니뇨 발생기간	SOI 자료를 사용한 빈도분석		RC1-6 자료를 사용한 빈도분석	
	Max(-SOI)	재현기간(년)	Max(-RC(1-6))	재현기간(년)
1940 년 - 1941년	2.6	12	1.4	12
1946 년 - 1947 년	1.8	3	0.7	3
1951 년 - 1952 년	1.4	2	0.8	3
1953 년 - 1954 년	1.5	2	0.6	2
1957 년 - 1958 년	2.3	7	0.6	2
1965 년 - 1966 년	1.8	3	1.1	6
1969 년 - 1970 년	1.6	2	0.7	3
1972 년 - 1973 년	2.0	4	1.2	7
1977 년 - 1978 년	3.5	27	1.0	4
1982 년 - 1983 년	4.6	133	2.3	107
1987 년 - 1988 년	1.2	2	1.1	5
1991 년 - 1992 년	3.4	19	1.4	10
1993 년 - 1994 년	1.8	3	1.0	4
1997 년 - 1998 년	3.5	44	1.8	27

표2에서 1940년, 1946년, 1951, 1953년, 1969년, 1993년 엘니뇨 사상에 대해서는 사용된 자료에 상관없이 비슷한 재현기간 추정되었으나, 나머지 엘니뇨 사상은 다르게 나타났다. 특히, 1977년 엘니뇨 사상은 SOI 자료(재현기간 27년)와 RC 1-6 자료(재현기간 4년)의 재현기간이 큰 차이로 다르게 나타났다. 이 경우는 1977년에 발생한 엘니뇨의 SOI 값이 -1.0과 0의 범위에서 대부분 발생하다가 78년 2월 갑자기 -3.5의 SOI 값이 발생한 까닭이다. 이와 같이 이례적으로 발생한 어느 한 달의 SOI 값이 엘니뇨 사상의 크기를 나타내기에는 문제가 있다. 이와 같은 경우에는 그림2에서와 같이 SSA를 통하여 자료를 필터링한 후 구성된 RC1-6으로 빈도해석하는 것이 좀더 타당한 것으로 나타났다. 따라서 RC1-6을 사용한 빈도해석을 살펴보면 지금까지 최대의 엘니뇨 발생 연도는 1982-83년으로 재현기간이 107년으로 나타났다. 두 번째 크기의 엘니뇨는 1997-98년에 발생되었으며 재현기간은 27년이다.

결론적으로 엘니뇨 사상 및 라니냐 사상에 대한 크기별 빈도해석에는 SOI 자료보다 SSA 기법을 사용하여 자료의 비조화성분을 제거한 RC 계열의 자료를 사용하는 것이 더욱 합리적인 것으로 나타났다.

5. 참 고 문 헌

- Barnett, T., N. Graham, M. Cane, S. Zeibak, S. Dolan, J. O'Brien, and D. Legler, 1988, On the Prediction of the El nino of 1986-87, *Science*, 241, 192-196.
- Broomhead, D. S., and G. P. King, 1986, Extracting Qualitative Dynamics from Experimental Data, *Phys. D Amsterdam*, 20, 217-236.
- Colebrook, J. M., 1978, Continuous Plankton Records: Zooplankton and Environment, North-East Atlantic and North Sea, 1948-1975, *Oceanol. Acta* 1: 9-23.
- Fraedrich, K., 1986, Estimating the Dimensions of Weather and Climate Attractors, *J. Atmos. Sci.*, 43, 419-432.
- Ghil, M. and R. Vautard, 1991, Interdecadal Oscillations and the Warming Trend in Global Temperature Time Series, *Nature*, 350, 324-327.
- Keppenne C. L., and M. Ghil, 1992, Adaptive Filtering and Prediction of the Southern Oscillation Index, *Journal of Geophysical Research*, vol. 97, No. D18, 20449 - 20454.
- Kiladis, G. N. and H. F. Diaz, 1989, Global Climatic Anomalies Associated with Extremes in the Southern Oscillation, *Journal of Climate*, 2(9), 1069-1090.
- Lall, U., Young-II Moon, and K. Bosworth., 1993, Kernel flood frequency estimators : bandwidth selection and kernel choice. *Water Resources Research* 29(4):1003-1015.
- Moon, Young-II and U. Lall, April 1996, Atmospheric Flow Indices and Interannual Great Salt Lake Variability, *Journal of Hydrologic Engineering*, 55-62.
- Moon, Young-II, Lall, U., and Bosworth, K., 1993, A comparison of tail probability estimators. *Journal of Hydrology* 151:343-363.
- Rasmusson, E. M., X. Wang, and C. F. Ropelewski, 1990, The Biennial Component of ENSO Variability, *J. Mar. Sys.*, 1, 71-96.
- Rosenblatt, M., 1956, Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27:832-837.
- Scott, D. W., 1992, *Multivariate Density Estimation*, John Willy Interscience, New York.
- Silverman, B.W., 1986, *Density estimation for statistics and data analysis*. Chapman and Hall, New York.
- Trenberth, K. E., 1997, The Definition of El nino, *Bulletin of the American Meteorological Society*, 78(12), 2771-2777.
- Vautard, R., and M. Ghil, 1989, Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time Series, *Phys. D, Amsterdam*, 35, 395-424.
- Vautard R., S. Yiou, and M. Ghil, 1992, Singular Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals, *Physica D*, 58, 95-126.